

# СИСТЕМНЫЙ АНАЛИЗ И ОБУЧАЮЩИЕ СИСТЕМЫ

## SYSTEMS ANALYSIS AND TRAINING SYSTEMS

---

УДК 004.048

**С. Г. Шульдова, А. И. Парамонов, Д. М. Карнаух, Н. В. Лапицкая**  
Белорусский государственный университет информатики и радиоэлектроники

### **АНАЛИЗ НАУЧНОЙ ДЕЯТЕЛЬНОСТИ КАФЕДРЫ В КОНТЕКСТЕ ПРИОРИТЕТНЫХ НАПРАВЛЕНИЙ ИССЛЕДОВАНИЙ И НАУЧНОГО СОТРУДНИЧЕСТВА**

В статье рассмотрены методы наукометрического анализа в контексте решаемых задач наукометрии. Предложен подход к решению задачи выявления научных сообществ и оценки научного взаимодействия сотрудников структурного подразделения (на примере кафедры программного обеспечения информационных технологий БГУИР). Выполнена постановка задачи и приведено ее формальное описание. Для апробации предложенного подхода получен набор данных из открытых профилей сотрудников Google Академии. После предварительной обработки, включающей дедупликацию и очистку данных, сформированы лексический корпус и терм-документная матрица, преобразованная в матрицу смежности, на основе которой построен граф соавторства. Рассмотрены основные параметры графа и выполнена кластеризация его вершин. Для кластеризации использован метод, оптимизирующий значение модулярности. Так как модулярность зависит от числа шагов перехода из одной вершины в другую, компьютерный эксперимент проведен при различных значениях параметра. В результате выявлены одиннадцать сообществ, соответствующих максимальному значению модулярности. Проанализированы некоторые характеристики сообществ: процент внешних авторов, продуктивность и цитируемость. Выработаны рекомендации по возможному использованию результатов анализа научной деятельности на кафедре. Обозначены пути дальнейших исследований.

**Ключевые слова:** наукометрия, метод, анализ, взаимодействие, граф соавторства, кластеризация, модулярность, сообщества.

**Для цитирования:** Шульдова С. Г., Парамонов А. И., Карнаух Д. М., Лапицкая Н. В. Анализ научной деятельности кафедры в контексте приоритетных направлений исследований и научного сотрудничества // Труды БГТУ. Сер. 3, Физико-математические науки и информатика. 2023. № 1 (266). С. 71–76. DOI: 10.52065/2025-6141-2023-266-1-12.

**S. G. Shuldova, A. I. Paramonov, D. M. Karnaukh, N. V. Lapitskaya**  
Belarusian State University of Informatics and Radioelectronics

### **ANALYSIS OF SCIENTIFIC ACTIVITIES OF THE DEPARTMENT IN THE CONTEXT OF PRIORITY DIRECTIONS OF RESEARCH AND SCIENTIFIC COOPERATION**

The article deals with the methods of scientometric analysis in the context of the tasks of scientometrics to be solved. An approach is proposed for solving the problem of identifying scientific communities and assessing the scientific interaction of employees of a structural subdivision (on the example of the Information Technology Software Department of BSUIR). The problem is formulated and its formal description is given. To test the proposed approach, a data set was obtained from public profiles of Google Academy employees. After pre-processing, including data deduplication and cleaning, a lexical corpus and a term-document matrix are formed, which are converted into an adjacency matrix, on the basis of which a co-authorship graph is built. The main parameters of the graph are considered and the clustering of its vertices is performed. For clustering, a method was used that optimizes the value of modularity. Since modularity depends on the number of transition steps from one vertex to another, a computer experiment was carried out for different values of the parameter. As a result, 11 communities were identified, corresponding to the maximum value of modularity. Some characteristics of communities are analyzed:

the percentage of external authors, productivity and citation. Recommendations on the use of information on scientific activities at the department have been developed. Ways for further research are outlined.

**Keywords:** scientometrics, method, analysis, interaction, co-authorship graph, clustering, modularity, communities.

**For citation:** Shuldova S. G., Paramonov A. I., Karnaukh D. M., Lapitskaya N. V. Analysis of scientific activities of the department in the context of priority directions of research and scientific cooperation. *Proceedings of BSTU, issue 3, Physics and Mathematics. Informatics*, 2023, no. 1 (266), pp. 71–76. DOI: 10.52065/2025-6141-2023-266-1-12 (In Russian).

**Введение.** В настоящее время для измерения продуктивности научной деятельности отдельных исследователей, научно-исследовательских организаций и университетов широко используются наукометрические методы, которые основаны на оценке конечных результатов научной деятельности: статей, монографий, методических разработок, пособий и книг. Данные методы также применяют для выявления актуальных и перспективных научных направлений, детектирования научных сообществ, предотвращения массового тиражирования работ, лишения ценности.

Среди методов наукометрического анализа, как правило, выделяют [1]:

- количественную оценку элементов документального информационного потока (ДИП);
- цитатный анализ;
- семантический анализ текстов научных публикаций.

Применение того или иного метода обуславливается решаемой с его помощью задачей [2], а реализация может осуществляться средствами различных технологий.

**Основная часть.** Методы количественной оценки ДИП обеспечивают решение таких задач, как распределение публикаций по типу издания, странам и учреждениям, подготовившим и издавшим научные документы; видам научных документов (книги, патенты и авторские свидетельства, публикации в журналах и т. д.) и носителей; количеству авторов; тематической направленности публикаций или журналов; числу библиографических ссылок в публикациях; «возрасту» публикаций; языковой принадлежности научных документов; объему документов и т. д.

Методы цитатного анализа позволяют определять вклад в мировую науку отдельных ученых и научных учреждений, выявлять структуру областей знания и даже прогнозировать их развитие. Библиографические ссылки являются своеобразными «индикаторами» информационных связей между научными документами и их авторами, необходимым условием научной этики.

Наряду с анализом цитирования при выявлении приоритетных направлений и детектировании научных сообществ часто используются методы семантического анализа метаданных и

полных текстов научных публикаций. В контент-анализе исследуется частотность употребления слов в метаданных и полных текстах и отдельно – ключевых слов, а также их совместная встречаемость в публикациях. Контент-анализ позволяет выявлять тенденции развития научных дисциплин и отдельных научных направлений, структурировать поток информации и выявлять определенные закономерности развития содержания научных публикаций, осуществлять прогнозные исследования, тематическую направленность отдельных групп публикаций.

Необходимые для исследования данные могут быть получены из наукометрических баз данных и систем цитирования, среди которых наиболее известными являются Web of Science, Scopus, Google Scholar и Российский индекс цитирования (РИНЦ).

В качестве индикаторов оценки научной деятельности используются следующие показатели [3]:

– количество публикаций: статьи, зачисленные в авторский профиль ученого или профиль научной организации. Это показатель научной производительности, по которому можно оценивать автора, группу авторов, организацию, государство, журнал;

– индекс цитирования (Science Citation Index, SCI): показатель «значимости» научной работы ученого или коллектива. Величина индекса цитирования определяется количеством ссылок на публикацию или фамилию автора в других источниках;

– индекс Хирша (h-index): наукометрический показатель, предложенный в 2005 г. Хорхе Хиршем из Калифорнийского университета в Сан-Диего. Индекс Хирша основан на учете числа публикаций исследователя и числа цитирований этих публикаций. Ученый имеет индекс h, если h есть наибольшее число статей, каждая из которых процитирована как минимум h раз каждая.

В настоящей работе рассматривается задача выявления научных сообществ и исследования их структуры в рамках структурного подразделения (кафедры) университета. В качестве исходных данных используются данные открытых профилей Google Scholar профессорско-преподавательского состава кафедры программного

обеспечения информационных технологий Белорусского государственного университета информатики и радиоэлектроники.

Для каждой статьи из открытого профиля сотрудника в Google Scholar были получены следующие характеристики: автор (группа авторов), название, вид публикации, цитируемость, издатель, год публикации, выпуск (номер для журнала), URL-адрес статьи, URL-адрес цитирования, номера страниц.

Таким образом, статья может быть представлена как кортеж:

$$Article_i = \langle Title_i, Keys_i, Authors_i, Type, Ref_i, SCI_i, URL_i, Publish, Year \rangle, \quad (1)$$

где  $Title_i$  – название,  $Keys_i$  – множество ключевых слов,  $Authors_i$  – множество авторов,  $Type$  – тип,  $Ref_i$  – аннотация (или  $Text_i$  – текст),  $SCI_i$  – цитируемость,  $URL_i$  – адрес цитирования,  $Publish$  – издание,  $Year$  – год публикации.

Автор, соответственно, – кортеж из таких характеристик, как фамилия, имя, индекс Хирша, научная степень, организация:

$$Author_k = \langle Surname_k, Name_k, h-index_k, Degree, Organization \rangle. \quad (2)$$

Тогда научное сообщество можно определить как сеть ученых, связанных отношениями соавторства и цитирования, а также общей тематикой статей.

$$Sch_k = \langle CAuthors, CKeys \rangle; \quad CAuthors \subset Authors, CKeys \subset Keys. \quad (3)$$

Решение данной задачи может быть получено на основе графов соавторства и цитирования, а также семантическом анализе текстов (аннотаций) публикаций выявленных сообществ.

Определим отношение цитирования как

$$Article_j \mathfrak{R}^r Article_i \equiv Article_j \text{ цитирует } Article_i,$$

тогда граф цитирования представляет собой ориентированный граф  $G^{ar} = (Articles, \mathfrak{R}^r)$ , в котором вершинами являются научные статьи, а направленными (ориентированными) дугами из множества  $\mathfrak{R}^{ar}$  – ссылки на статьи внутри другой статьи, то есть  $Articles = \{ar_1, ar_2, \dots, ar_n\}$  соответствуют статьям, а дуги – отношениям цитирования  $\mathfrak{R}^{ar} = Articles \times Articles$ ,  $r^{ar} = \langle ar_j, ar_i \rangle \in \mathfrak{R}^{ar}$ ,  $i \neq j$ , так как самоцитирование исключено. Также необходимо учесть, что  $ar_j$  может цитировать  $ar_i$  только один раз, и если  $ar_j$  цитирует  $ar_i$ , то  $ar_i$  не может цитировать  $ar_j$ , то есть петель, кратных ребер и циклов в графе нет, а вес ребра, соединяющего цитируемую и цитирующие статьи,  $f(r^{ar}) = 1$ .

Для авторов определим отношение соавторства: авторы являются соавторами, если существует

статья, множеству авторов которой принадлежит оба автора:

$$Author_k \mathfrak{R}^{au} Author_l, (\exists Article_i \in Articles); \\ Author_k \in Authors_i, Author_l \in Authors_i.$$

Граф соавторства представляет собой неориентированный граф  $G^{au} = (Authors, \mathfrak{R}^{au})$ , без петель и кратных ребер, вершины  $Authors = \{au_1, au_2, \dots, au_m\}$  которого соответствуют авторам, ребра – отношениям соавторства  $r^{au} = \langle au_k, au_l \rangle \in \mathfrak{R}^{au}$ ,  $k \neq l$ . В простейшем случае, если количество совместных статей и вклад каждого соавтора не учитываются, вес ребра равен 1. Такой граф отражает только наличие сотрудничества [4] и задается матрицей смежности  $U$ :

$$u_{k,l} = \begin{cases} 1, \text{ если } au_k \mathfrak{R}^{au} au_l, \\ 0, \text{ в противном случае.} \end{cases}$$

Матрица смежности графа соавторства получена умножением  $A \cdot A^T$ , где  $A$  – матрица авторства:

$$A_{k,i} = \begin{cases} 1, \text{ если } au_k \in Authors_i, \\ 0, \text{ в противном случае.} \end{cases}$$

Граф цитирования авторов  $G^{ru} = (Authors, \mathfrak{R}^{ru})$  строится на основе данных о цитировании без самоцитирования авторами, то есть

$$Author_k \mathfrak{R}^{ru} Author_l, \text{ если } \mathfrak{R}^{ru} \subset \mathfrak{R}^{ar}; \\ Authors_i \cap Authors_j = \emptyset,$$

$i \neq j, k \neq l$ , и представляет собой ориентированный граф без петель.

Построение и исследование графов выполнено в среде R Studio средствами статистического языка программирования R [5].

Препроцессинг данных являлся самым трудозатратным процессом анализа и состоял из следующих шагов:

- 1) удаление дубликатов;
- 2) удаление строк с пропущенными значениями в названии либо авторах (при наличии);
- 3) транслитерация фамилий авторов;
- 4) очистка фамилий авторов и приведение к виду ИО\_ФАМИЛИЯ;
- 5) очистка данных изданий и конструирование характеристики принадлежности издания перечню изданий Высшей аттестационной комиссии для публикации результатов диссертационных исследований.

Построение графа соавторства осуществлялось в несколько этапов и включало в себя создание лексического корпуса и терм-документной матрицы, которая представляет собой матрицу авторства, преобразование ее в матрицу смежности графа и непосредственно визуализацию графа.

Полученный граф соавторства состоит из 386 вершин и 695 ребер. Плотность сети определяется как отношение числа имеющихся в графе связей к максимально возможному и указывает

на степень влияния авторов друг на друга, для исследуемого графа плотность достаточно мала – 0,0066. Транзитивность отражает отношение «соавтор моего соавтора – мой соавтор» и равна 0,21. Степени вершин показывают количество соавторов данного автора, максимальная степень равна 58, средняя степень вершин – 3,6.

Для детектирования сообществ применялся метод WalkTrap, использующий в качестве критерия модулярность – меру, которая равна доле ребер от общего числа ребер, которые попадают в данные группы, минус ожидаемая доля ребер, которые попали бы в те же группы, если бы они были распределены случайно. Объект может переместиться из вершины  $i$  в вершину  $j$  с вероятностью  $P_{ij} = u_{ij} / d_j$ , где  $u_{ij}$  – элемент матрицы смежности графа,  $d_j$  – степень  $j$ -й вершины. То есть на каждом шаге равновероятно выбирается «сосед» вершины  $i$ . Таким образом определяется матрица переходов  $P$  случайного блуждания. Она примечательна тем, что ее степени являются вероятностями перехода из одной вершины в другую за соответствующее число шагов: вероятность перехода из  $i$  в  $j$  за  $t$  шагов равна  $(P^t)_{ij}$ . Эксперименты были проведены с различным значением количества шагов  $t$ :  $3 \leq t \leq 8$  [6]. Максимальное значение модулярности было получено при  $t = 8$ . Таким образом, было выявлено 11 сообществ с модулярностью 0,86, что является более чем приемлемым результатом.

Граф с сообществами показан на рис. 1. Каждая вершина – это автор, ребра – совместная публикация. Авторы одного сообщества раскрашены в один цвет, граф не взвешенный, поэтому длина ребер произвольная для корректного отображения. Граф позволяет наглядно сравнить размеры сообществ и оценить их структуру и связанность членов.

Необходимо отметить, что сфера научных интересов одной из групп, которая является изолированной, отлична от остальных, так как ее авторы не имеют общих публикаций с другими сотрудниками кафедры. В ее состав входит только один представитель кафедры, проработавший чуть более года до проведения настоящего исследования.

Исходя из состава выявленных сообществ был определен процент сотрудников кафедры в составе каждой группы. Максимальное значение 34,3% получено для сообщества, включающего сотрудников, работающих на кафедре продолжительное время, в то время как минимальное 1,96% – для изолированного сообщества. Стаж работы на кафедре существенно влияет на количество совместных работ. Небольшие сообщества включают, как правило, одного представителя кафедры, уже имеющего совместные публикации с коллегами, но основной объем работ выполнившего на предыдущих местах работы.

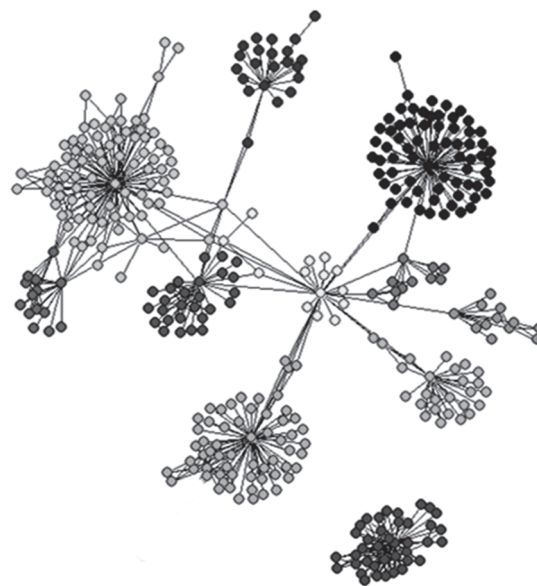


Рис. 1. Граф соавторства с сообществами

Для оценки продуктивности авторов сообществ сопоставлены количество авторов сообщества и количество статей (рис. 2), ими опубликованных, отсортированных по модулярности сообществ. Несмотря на то, что количество публикаций, безусловно, зависит от количества авторов, эта зависимость не линейна. На количество публикаций влияют также такие факторы, как остепенность и обучение в аспирантуре, получение научного звания, которые стимулируют публикационную деятельность.

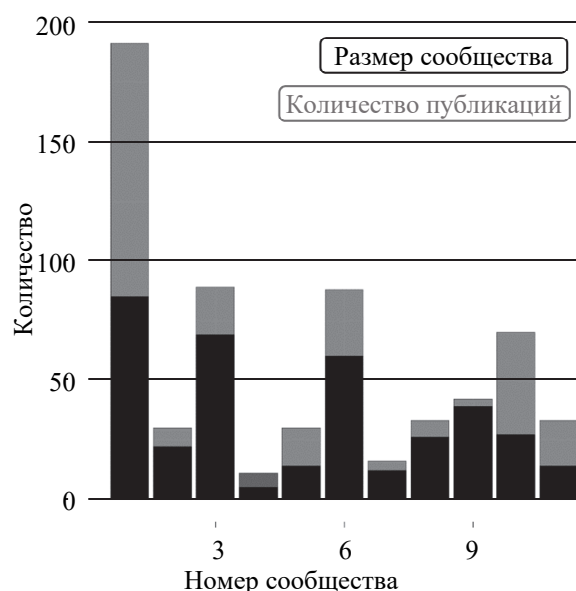


Рис. 2. Соотношение числа авторов к числу статей сообщества

Для оценки значимости сообщества проанализированы количество публикаций авторов и показатели цитируемости.

## Средняя цитируемость научных сообществ

Номер сообщества	1	2	3	4	5	6	7	8	9	10	11
Средняя цитируемость	7,03	0,93	1,47	2,6	2,20	0,94	2,00	0,54	2,03	1,78	5,43

В таблицу сведены данные о средней цитируемости сообществ.

Группы с наибольшей цитируемостью: в первой группе 85 авторов опубликовали 191 статью, в одиннадцатой группе 14 авторов – 33 статьи, в четвертой группе 11 авторов – 5 статей. В третьей и шестой группах, достаточно многочисленных, низкая цитируемость обусловлена наличием большого количества малоцитируемых или нецитируемых статей.

Также необходимо учитывать возраст публикации, с его ростом цитируемость увеличивается. Этим также можно обосновать низкую цитируемость в восьмой группе. Это публикации последних лет молодых сотрудников. Также необходимо заметить, что наличие одной высокоцитируемой публикации, например учебника, существенно влияет на результат (ситуация в одиннадцатой группе).

С учетом полученных результатов исследований представляется целесообразным анализ состава авторов выявленных сообществ, их локаций и аффилиации, а также анализ публикаций

сообщества, включая семантический анализ текстов, ранжирование по типу публикации и изданию, географические и временные распределения публикаций.

**Заключение.** Предложенный в статье подход к оценке научной деятельности сотрудников и степени их сотрудничества позволил выделить тематические сообщества (группы по научным интересам) и исследовать их структуру. В результате компьютерного эксперимента по исследованию научных связей сотрудников кафедры на основе их публикаций в базе Google Академии выявлено на кафедре 11 сообществ при уровне модулярности 0,86. Также определено, что ни одно из выделенных сообществ не состоит полностью только из сотрудников кафедры. Это показывает высокий уровень внешнего сотрудничества кафедры с учеными из других структур. В дальнейшем планируется определить «круги» сотрудничества и выявить «научных партнеров», а также исследовать географические и временные распределения, проанжировать статьи по типам и изданиям.

## Список литературы

1. Тронин В. Г. Оценка результатов научно-исследовательской работы и наукометрия. Ульяновск: УлГТУ, 2019. 136 с.
2. Руководство по наукометрии: индикаторы развития науки и технологии / М. А. Акоев [и др.] Екатеринбург: Урал, 2014. 250 с.
3. Войнилов Ю. Л. Индикаторы науки // Статистический сборник НИУ ВШЭ. 2017. URL: <https://www.hse.ru/primarydata/in2017> (дата обращения: 25.09.2022).
4. Бредихин С. В., Щербакова Н. Г. Коллективное авторство научных // Проблемы оптимизации сложных систем – 2020. URL: <https://cyberleninka.ru/article/n/kollektivnoe-avtorstvo-nauchnyh-publikatsiy1/viewer> (дата обращения: 25.12.2022).
5. IGraph package // R documentation. URL: <https://www.rdocumentation.org/packages/igraph/versions/1.3.5> (дата обращения: 25.09.2022).
6. Pons P. Computing communities in large networks using random walks // Physical Review E., 2005. URL: <http://arxiv.org/abs/physics/0512106v1> (дата обращения: 25.09.2022).

## References

1. Tronin V. G. *Otsenka rezultatov nauchno-issledovatel'skoy raboty i naukometriya* [Evaluation of the results of research work and scientometrics]. Ulyanovsk, UIGTU Publ., 2019. 136 p. (In Russian).
2. Akoev M. A., Markusova V. A., Moskaleva O. V., Pisyakov V. V. *Rukovodstvo po naukometrii: indikatory razvitiya nauki i tekhnologii* [Guide to scientometrics: indicators of the development of science and technology]. Ekaterinburg, Ural Publ., 2014. 250 p. (In Russian).
3. Voynilov Y. L. Science Indicators. *Statisticheskij sbornik NIU VSHE* [Statistical compendium of the National Research University Higher School of Economics], 2017. Available at: <https://www.hse.ru/primarydata/in2017> (accessed 25.09.2022) (In Russian).
4. Bredikhin S. V., Shcherbakova N. G. Collective authorship of scientific publications. *Problemy optimizatsii slozhnykh sistem – 2020* [Problems of optimization of complex systems – 2020]. Available at: <https://cyberleninka.ru/arti-cle/n/kollektivnoe-avtorstvo-nauchnyh-publikatsiy1/viewer> (accessed 25.12.2022).

5. IGraph package. *R documentation*. Available at: <https://www.rdocumentation.org/packages/igraph/versions/1.3.5> (accessed 25.09.2022).

6. Pons P. Computing communities in large networks using random walks. *Physical Review E.*, 2005. Available at: <http://arxiv.org/abs/physics/0512106v1> (accessed 25.09.2022).

### Информация об авторах

**Шульдова Светлана Георгиевна** – кандидат технических наук, доцент, доцент кафедры программного обеспечения информационных технологий. Белорусский государственный университет информатики и радиоэлектроники (220005, г. Минск, ул. Гикало, 9, Республика Беларусь). E-mail: [shsg@bsuir.by](mailto:shsg@bsuir.by).

**Парамонов Антон Иванович** – кандидат технических наук, доцент, заведующий кафедрой информационных систем и технологий. Институт информационных технологий Белорусского государственного университета информатики и радиоэлектроники (220037, г. Минск, ул. Козлова, 28, Республика Беларусь). E-mail: [a.paramonov@bsuir.by](mailto:a.paramonov@bsuir.by).

**Карнаух Дарья Михайловна** – магистрант, ассистент кафедры информационных систем и технологий. Институт информационных технологий Белорусского государственного университета информатики и радиоэлектроники (220037, г. Минск, ул. Козлова, 28, Республика Беларусь). E-mail: [d.karnaukh@bsuir.by](mailto:d.karnaukh@bsuir.by).

**Лапицкая Наталья Владимировна** – кандидат технических наук, доцент, заведующий кафедрой программного обеспечения информационных технологий. Белорусский государственный университет информатики и радиоэлектроники (220005, г. Минск, ул. Гикало, 9, Республика Беларусь). E-mail: [lapan@bsuir.by](mailto:lapan@bsuir.by).

### Information about the authors

**Shuldova Svetlana Georgievna** – PhD (Engineering), Associate Professor, Assistant Professor, the Information Technology Software Department. Belarusian State University of Informatics and Radioelectronics (9, Gikalo str., 220005, Minsk, Republic of Belarus). E-mail: [shsg@bsuir.by](mailto:shsg@bsuir.by)

**Paramonov Anton Ivanovich** – PhD (Engineering), Associate Professor, Head of the Information Systems and Technologies Department. Institute of Information Technology of the Belarusian State University of Informatics and Radioelectronics (28, Kozlova str., 220037, Minsk, Republic of Belarus). E-mail: [a.paramonov@bsuir.by](mailto:a.paramonov@bsuir.by)

**Karnaukh Daria Mikhailovna** – Master's degree student, Assistant Lecturer, the Information Systems and Technologies Department. Institute of Information Technology of the Belarusian State University of Informatics and Radioelectronics (28, Kozlova str., 220037, Minsk, Republic of Belarus). E-mail: [d.karnaukh@bsuir.by](mailto:d.karnaukh@bsuir.by)

**Lapitskaya Natalya Vladimirovna** – PhD (Engineering), Associate Professor, Head of the Information Technology Software Department. Belarusian State University of Informatics and Radioelectronics (9, Gikalo str., 220005, Minsk, Republic of Belarus). E-mail: [lapan@bsuir.by](mailto:lapan@bsuir.by)

*Поступила после доработки 18.01.2023*